

## Google Scholar, Sci-Hub and LibGen: Could they be our New Partners?

Louis Houle  
*McGill University (Canada)*

---

Louis Houle, "Google Scholar, Sci-Hub and LibGen: Could they be our New Partners?." *Proceedings of the IATUL Conferences*. Paper 6.  
<https://docs.lib.purdue.edu/iatul/2017/partnership/6>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries.  
Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.



**Library Genesis<sup>1M</sup>**

**Could they be our new partners?**

**Louis Houle – Director Collections  
McGill University Library**

11.2

# Goals

## ➤ Google Scholar (GS)

- ❖ Discovery/Searching tool
- ❖ How much content is there?
  - discipline/subject/database
  - Stability over time
- ❖ Currency of information
- ❖ Alternative to subscription-based abstracting and indexing services?



# Goals

## ➤ Sci-Hub & LibGen

- ❖ Full-Text Retrieving tools

- ❖ Content

- How much? Reliability/Stability?

- ❖ “New” Open Access tools?

- Replacing

- IR, Social Media (Academia, Mendeley, ResearchGate, etc.), P2P networks?

- ❖ Currency of information

- ❖ Alternatives to subscription-based peer review journals?



# ➤ GS



## ❖ Launched November 2004 (12 years ago)

- Scholarly search engine for both OA and non-OA articles
- Citation analysis tool
- Connects to library collections/catalogues
- Content Coverage: (<https://scholar.google.com/intl/en/scholar/about.html>)
  - Journal and conference papers
  - theses and dissertations
  - academic books
  - pre-prints
  - abstracts
  - technical reports
  - other scholarly literature from all broad areas of research
  - “works from a wide variety of academic publishers, professional societies and university repositories, as well as scholarly articles available anywhere across the web. GS also includes court opinions and patents.”



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Library Genesis<sup>1M</sup>

## ❖ LibGen

- Server based in Russia
- Over 52M articles
- 95% of its content is made of educational items such as scientific articles, books and textbooks
- 5% of recreational items like fiction books and comics
- Feeding LibGen:
  - “biblioleaks” (few isolated but massive addition of items)
  - Crowdsourcing (building a distributed collection of items shareable among users) – channels like Sci-Hub & Reddit
  - Principle: “Take and Give”



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Sci-Hub



## ❖ Sci-Hub

- Launched April 16, 2011 (6 years ago)
- Founder: Alexandra Elbakyan (Kazakhstan)
- Over 62M academic papers and articles
- Getting access to articles:
  - Credentials (IDs/passwords) from educational institutions
  - Anonymous academics giving their credentials
  - phishing
- Collaborating with LibGen since 2013
  - If the searched paper in Sci-Hub is available in LibGen :
    - then it will be made available in Sci-Hub without using any credentials
  - If paper not available in LibGen :
  - Sci-Hub will download it and store it in LibGen for any future requests



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Methodology

- Project from September 2016 to August 2017
- 55 databases (all disciplines)
  - Random sampling – peer review articles
    - Research Randomizer (<https://www.randomizer.org/>)
  - Total of 2,750 samples
    - 50 samples/database
- Searching the samples
  - Done quarterly during the year
    - Oct 2016, Jan 2017, Apr 2017 & Jul 2017
  - Google Scholar, Sci-Hub & LibGen
  - Article title, DOI, author, Journal title, ISSN



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service



# Methodology

- Currency of information in GS, Sci-Hub & LibGen
  - ❖ Searching research articles only
    - Nature: New issues and Nature Advance Online Publication (AOP)
    - Science: New issues and Science First Release
  - ❖ Searching via DOI and article title
  - ❖ In GS, Sci-Hub, LibGen and Google
  - ❖ Daily for a year



# List of Databases

Database	Subject from McGill	Platform
Art Full Text 1984 -	Arts	EBSCO/Wilson
ABI Inform Global 1905-	Business/Management	ProQuest
Business Source Complete	Business/Management	EBSCO
Canadian Business & Current Affairs Database (CBCA) 1971-	Business/Management	ProQuest
Econlit 1886-	Business/Management	EBSCO
Education Full text	Education	EBSCO/Wilson
ERIC 1966 -	Education	EBSCO
Library & Information Science Abstracts (LISA) 1969 -	Education	ProQuest
Library Literature & Information Science Full Text 1980-	Education	EBSCO/Wilson
ATLA Religion Database with ATLASerials 1908-	Humanities	EBSCO
British Humanities (BHI) Index 1962 -	Humanities	ProQuest
Communication Abstracts 1972 -	Humanities	EBSCO
Historical Abstracts 1955 -	Humanities	EBSCO
MLA International Bibliography 1962-	Humanities	ProQuest
PAIS Index 1914-	Humanities	ProQuest
Philosopher's Index 1940 -	Humanities	ProQuest
Periodicals Archive Online	Humanities/social Sciences	ProQuest
Francis 1972-	Humanities/Social Sciences (foreign language)	EBSCO

# List of Databases

Database	Subject from McGill	Platform
Index to Legal Periodicals & Books Full Text	Law	EBSCO/Wilson
Index to Foreign Legal Periodicals 1985-	Law (foreign language)	Hein Online
Music Periodicals Database 1874 -	Music	ProQuest
RILM Abstracts of Music Literature 1800-	Music	EBSCO
Linguistics and Language Behavior Abstracts (LLBA) 1973 -	Social Sciences	ProQuest
PsycINFO 1806 -	Social Sciences	Ovid
Social Services Abstracts 1979 -	Social Sciences	ProQuest
Social Work Abstracts 1968 -	Social Sciences	Ovid
SocINDEX with Full Text 1908 -	Social Sciences	EBSCO
Sociological Abstracts 1952 -	Social Sciences	ProQuest

# List of Databases

Database	Subject from McGill	Platform
Agricola 1970 -	Agriculture/Environment	Ovid
CAB Abstracts + Archive 1910 -	Agriculture/Environment	Ovid
Compendex 1884 -	Engineering	EI Village
Inspec 1896 -	Engineering/Science	EI Village
BIOSIS Previews 1969 -	Health/Biological Sciences	Ovid
CINAHL Plus with Full Text 1937 -	Health/Biological Sciences	EBSCO
Embase Classic + Embase 1947-	Health/Biological Sciences	Ovid
Medline 1946 -	Health/Biological Sciences	Ovid
PubMed 1946 -	Health/Biological Sciences	NIH
Zoological Record 1864 -	Health/Biological Sciences	Ovid
Applied Science & Technology Full Text 1983 -	Science	EBSCO/Wilson
GeoRef 1693 -	Science	ProQuest
MathSciNet 1800-	Science	AMS
Meteorological & Geostrophysical Abstracts 1974 -	Science	ProQuest
SciFinder Scholar 1907 -	Science	ACS

# List of Databases

Database	Subject from McGill	Platform
Academic Search Complete	Multidisciplinary	EBSCO
Expanded Academic ASAP	Multidisciplinary	Gale
OmniFile Full Text Mega	Multidisciplinary	EBSCO/Wilson
ProQuest Research Library	Multidisciplinary	ProQuest
Sage	Multidisciplinary	Sage
ScienceDirect	Multidisciplinary	Elsevier
Scopus	Multidisciplinary	Elsevier
SPORTDiscus	Multidisciplinary	EBSCO
Springer	Multidisciplinary	Springer-Nature
Taylor & Francis	Multidisciplinary	Taylor & Francis
Web of Science	Multidisciplinary	Thomson Reuters
Wiley	Multidisciplinary	Wiley

# 2,750 Samples

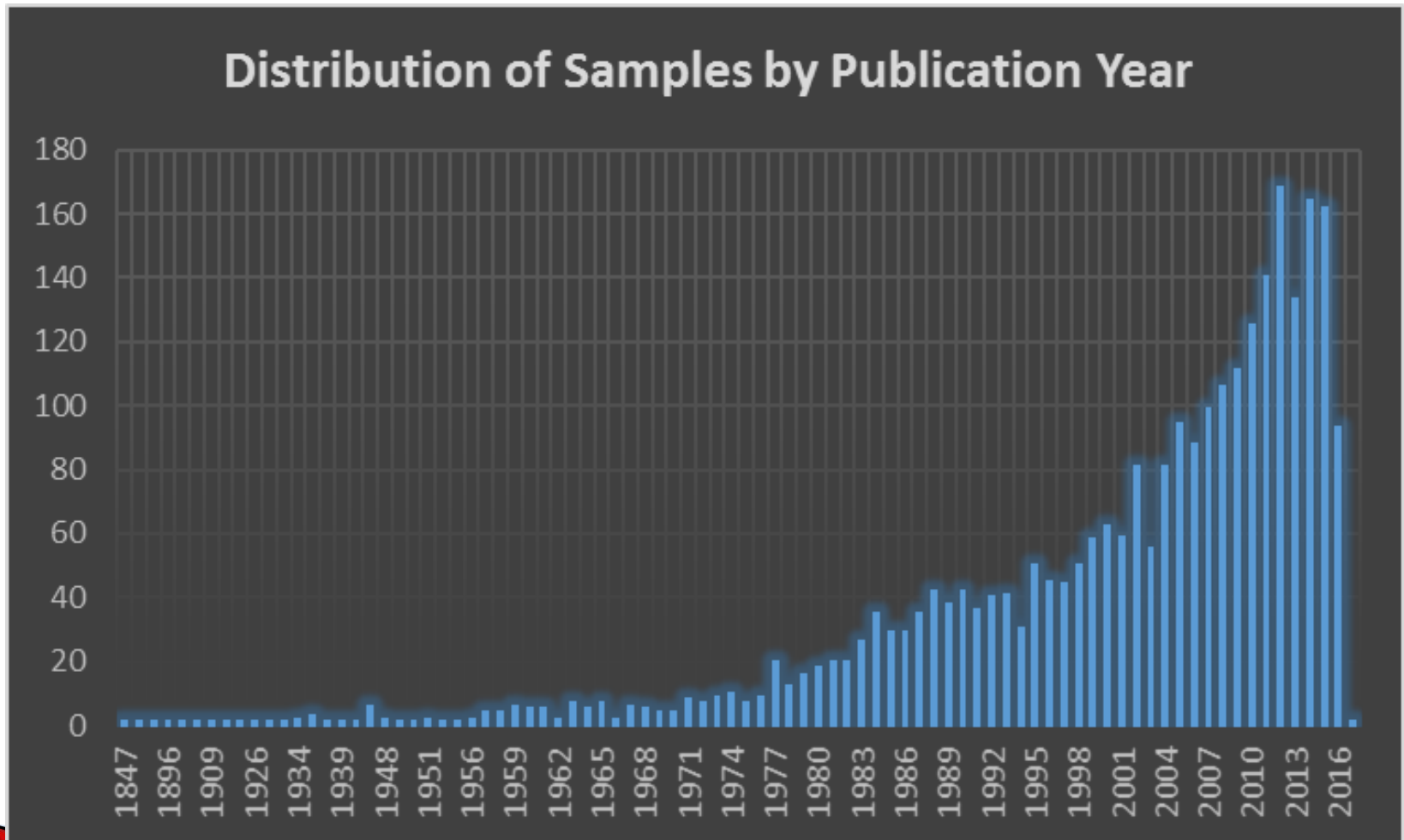
## ➤ Publication Years

- ❖ 86 different years
- ❖ 1847 - 2017
- ❖ 16 decades represented
- ❖ Complete decades
  - '60s
  - '70s
  - '80s
  - '90s
  - '00s
  - 2010-17



# 2,750 Samples

PY between 1847 to 2017



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# 2,750 Samples

## Rank of Top 10 publishers with # of samples

Publisher	Rank	# samples	% of total samples
Taylor & Francis	1	326	11.85
Elsevier	2	262	9.53
Wiley	3	243	8.84
Springer	4	206	7.49
Sage	5	166	6.04
Oxford	6	66	2.40
Cambridge	7	40	1.45
MCB-Emerald	8	29	1.05
IEEE (Institute of Electrical and Electronics Engineers)	9	25	0.91
BMC (BioMed Central)	10	21	0.76
Brill	10	21	0.76
		<b>1406</b>	<b>51.05</b>



**McGill**

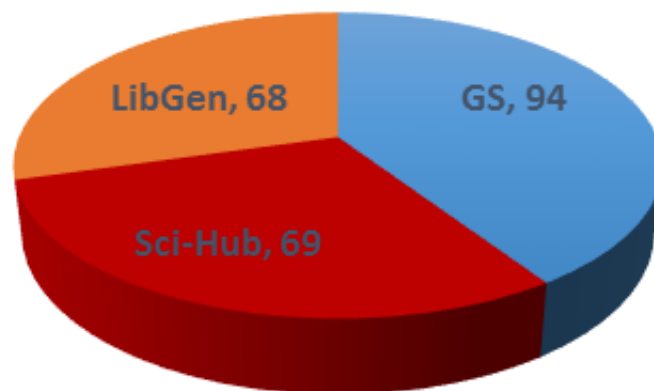
**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service



# Overall % of retrieved samples in GS, Sci-Hub & LibGen (Apr 2017)

Overall % of Retrieved Samples  
in GS, Sci-Hub & LibGen in Apr 2017



■ GS ■ Sci-Hub ■ LibGen



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Searched & Retrieved (Apr 2017)

- GS (2598)
  - ❖ 1660 by title (99.6%)
  - ❖ 11 by DOI (0.4%)
- Sci-Hub (1891)
  - ❖ 1661 by title (87.8%)
  - ❖ 224 by DOI (11.8%)
  - ❖ 6 by PMCID (0.3%)
- LibGen (1869)
  - ❖ 1670 by title (89.4%)
  - ❖ 196 by DOI (10.5%)
  - ❖ 3 (PMCID & journal title) (0.2%)



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Library Genesis: Scientific Articles $\lim_{n \rightarrow \infty}$

10.1007/s10551-008-9727-y > > > > Search!

Search in DOI or PMID (e.g. 10.1063/1.529338 or 16387140) OR Author+Article, if file not found - redirect to [sci-hub.bz](http://sci-hub.bz) ☒ Redirect

found - redirect to [sci-hub.bz](http://sci-hub.bz)

DOI	Author(s)	Article	DOI owner	Journal	Issue	ISSN	Size, kB	Other
DOI: 10.1007/s10551-008-9727-y								
Links: <a href="#">Libgen</a> <a href="#">Sci-Hub Moscow</a> <a href="#">Sci-Hub Ocean</a> <a href="#">Sci-Hub Cyber</a> <a href="#">BookSC</a> <a href="#">Libgen.net</a> <a href="#">Torrents</a>	Yingyan Wang	Examination on Philosophy-Based Management of Contemporary Japanese Corporations: Philosophy, Value Orientation and Performance	Springer- Verlag	<a href="#">Journal of Business Ethics</a>	Year:2009 Month:03 Day: Volume:85 Issue:1 First page:1 Last page:12	0167-4544(p) 1573-0697(e)	147kb	<a href="#">MD5</a> <a href="#">Errors</a> <a href="#">Edit</a> <a href="#">BibTeX</a> <a href="#">dx.doi.org</a> <a href="#">Abstract</a>



# Overall range of retrieval rates (%) (Apr 2017)

## ➤ Distribution over all 55 databases:

GS 58% - 100%

Sci-Hub 20% - 100%

LibGen 20% - 100%



# Lowest retrieval rates (%) by database (Apr 2017)

## GS Retrieval Rate

58%

76%

## Database

RILM Abstracts of Music Literature 1800-

Index to Foreign Legal Periodicals 1985-

## Sci-Hub Retrieval Rate

20%

28%

28%

32%

## Database

Index to Legal Periodicals & Books Full Text

Index to Foreign Legal Periodicals 1985-

RILM Abstracts of Music Literature 1800-

Canadian Business & Current Affairs Database (CBCA) 1971-

## LibGen Retrieval Rate

20%

28%

28%

30%

## Database

Index to Legal Periodicals & Books Full Text

Index to Foreign Legal Periodicals 1985 -

RILM Abstracts of Music Literature 1800 -

Canadian Business & Current Affairs Database (CBCA) 1971-



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Highest retrieval rates (%) by database (Apr 2017)

## GS Retrieval Rate

100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%  
100%

## Database

Agricola 1970 -  
Applied Science & Technology Full Text 1983 -  
Business Source Complete  
CINAHL Plus with Full Text 1937 -  
Econlit 1886-  
ERIC 1966 -  
Medline 1946 -  
ProQuest Research Library  
ScienceDirect  
SciFinder Scholar 1907 -  
Social Services Abstracts 1979 -  
Springer  
Taylor & Francis  
Web of Science  
Wiley

## Sci-Hub Retrieval Rate

100%  
100%  
100%  
96%  
96%  
94%  
90%

## Database

ScienceDirect  
Springer  
Wiley  
British Humanities (BHI) Index 1962 -  
Taylor & Francis  
Sage  
Meteorological & Geostrophysical Abstracts 1974 -

## LibGen Retrieval Rate

100%  
100%  
100%  
96%  
96%  
94%  
90%

## Database

ScienceDirect  
Springer  
Wiley  
British Humanities (BHI) Index 1962 -  
Taylor & Francis  
Sage  
Meteorological & Geostrophysical Abstracts 1974 -



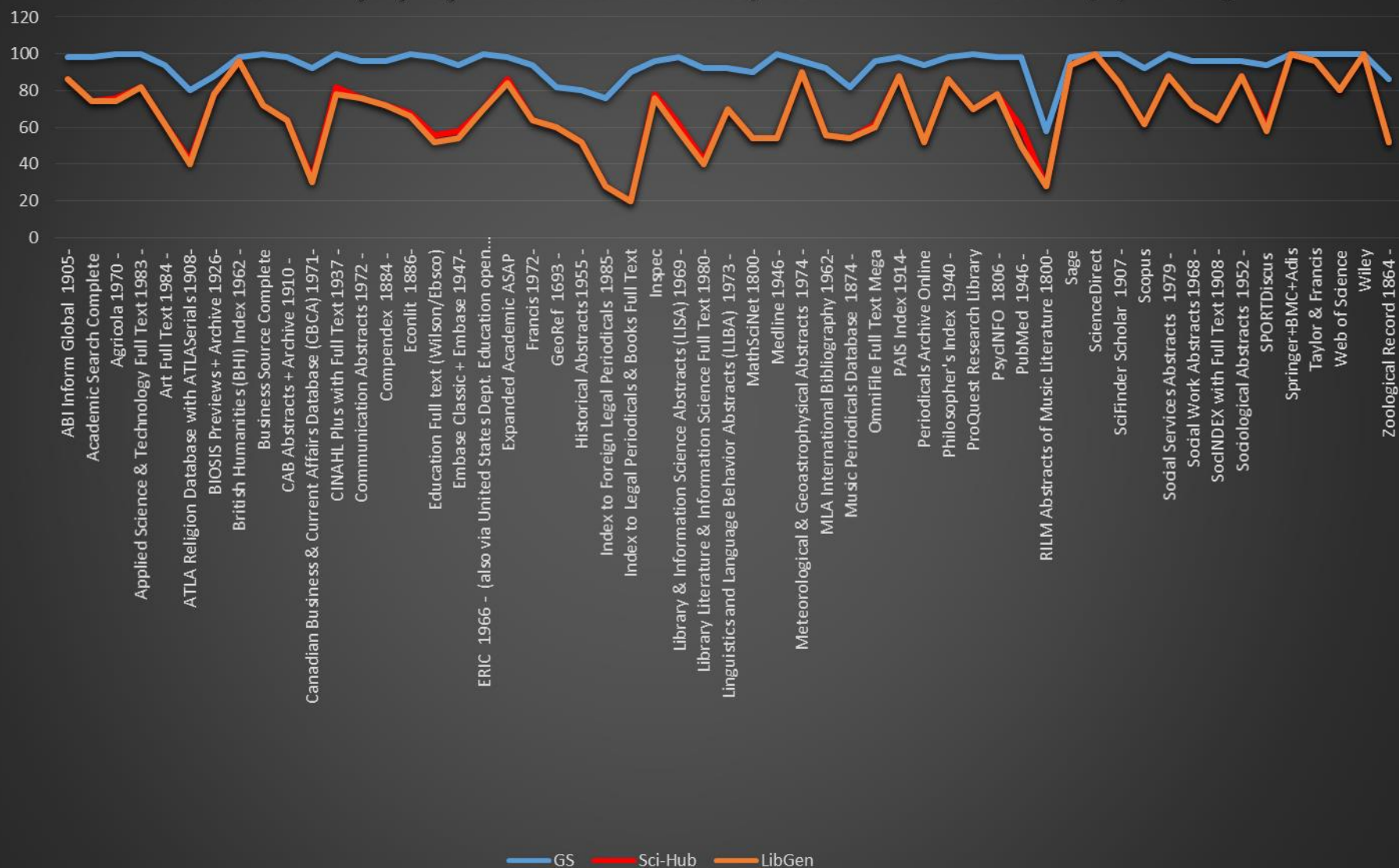
**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service



# Retrieval Rates (%) by database for GS, Sci-Hub & LibGen (Apr 2017)



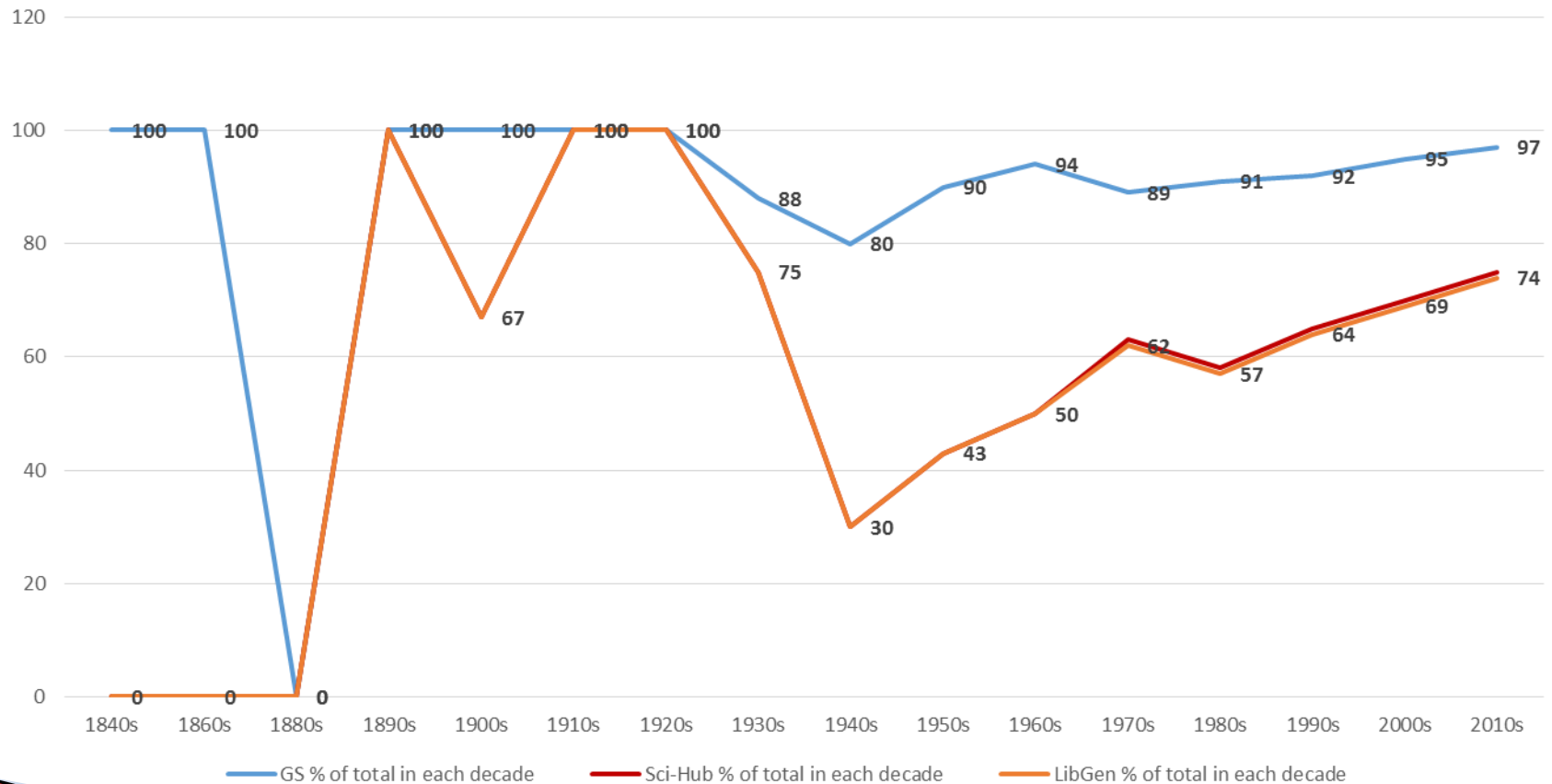
**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Distribution Rates by Decades GS, Sci-Hub & LibGen (Apr 2017)

Overall Distribution of Retrieval Rates by Decades for GS, Sci-Hub & LibGen (Apr 2017)



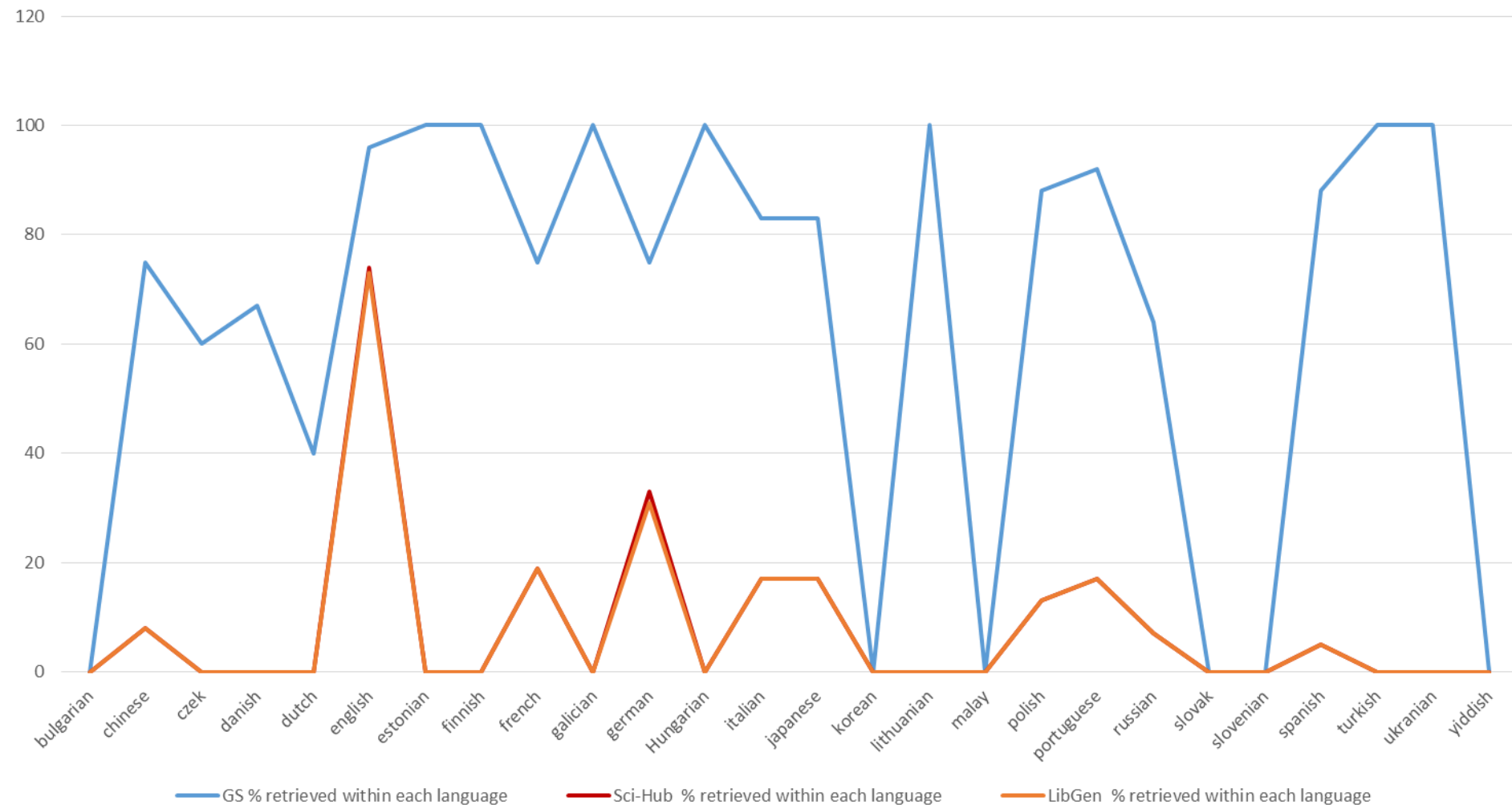
**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service



# Retrieval Rates (%) by Language for GS, Sci-Hub & LibGen (Apr 2017)

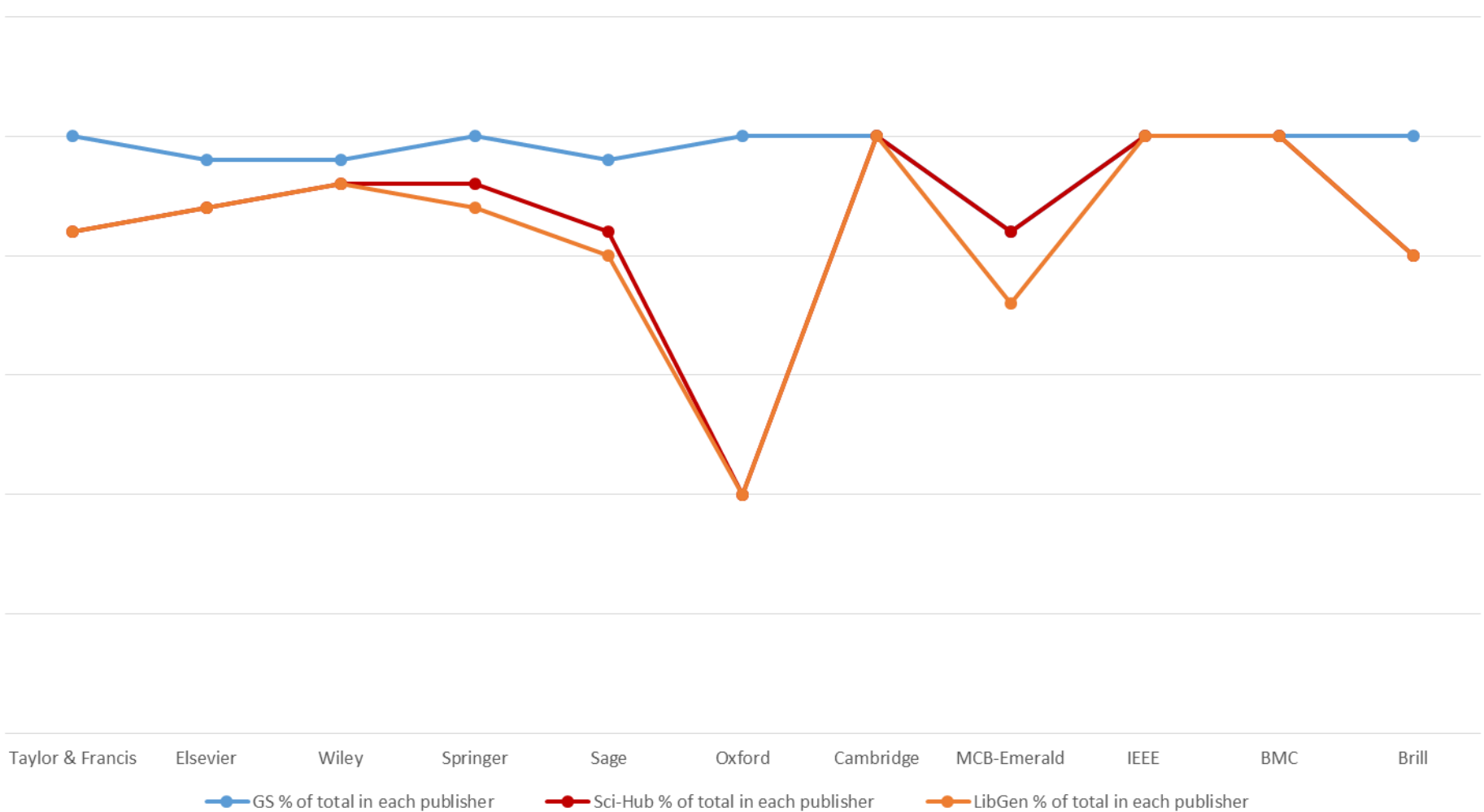


**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

Overall distribution of retrieval rates (%) by Publisher (Top 10 only) - Apr 2017



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Searching/Linking errors

## Sci-Hub & LibGen

### ➤ Sci-Hub

- ❖ Will bring different article from the searched one
- ❖ Same author & article title but different journal title/year
- ❖ Linking to web site where you need to buy article

### ➤ LibGen

#### ❖ Searching:

- Reaching max # sessions (40)
- Missing parts of title or typos

#### ❖ Retrieving

- Only by DOI or title since no other info in record
- Same author & article title but different journal title/year
- Showing wrong title on record but getting good one from the Sci-Hub link



**McGill**

**Library  
Bibliothèque**

Information ♦ Innovation ♦ Service

# Library Genesis: Scientific Articles $\lim_{n \rightarrow \infty}$

10.1017/S1047759400003433

Search in DOI or PMID (e.g. 10.1063/1.529338 or 16387140) OR Author+Article, if file not

Journal Title or ISSN

Volume or Year

Issue

Pages

Search!

☒ Redirect

found - redirect to [sci-hub.bz](http://sci-hub.bz)

DOI	Author(s)	Article	DOI owner	Journal	Issue	ISSN	Size, kB	Other
<b>DOI:</b> 10.1017/S1047759400003433								
<b>Links:</b> <a href="#">Libgen</a>					Year:			
<a href="#">Sci-Hub Moscow</a>					Month:			<a href="#">MD5</a>
<a href="#">Sci-Hub Ocean</a>			Cambridge		Day:			<a href="#">Errors</a>
<a href="#">Sci-Hub Cyber</a>			University		Volume:		970kb	<a href="#">Edit</a>
<a href="#">BookSC</a>			Press		Issue:			<a href="#">BibTeX</a>
<a href="#">Libgen.net</a>					First page:			<a href="http://dx.doi.org">dx.doi.org</a>
<a href="#">Torrents</a>					Last page:			



# Currency of Information (Sep. 2016 to May 2017)

## ➤ Nature

### ❖ Nature new issues (research articles only)

- Released every Thursday but in fact coming out Wednesday night
- 15 articles only (many published as AOP)
- All found by DOI except for 1 by title in GS
- GS: 73% 4-5 days. Range from a few hours to 6 days
- Sci-Hub: 100% few hours
- LibGen: 100% few hours
- Google: 100% few hours

# Currency of Information (Sep. 2016 to May 2017)

## ➤ Nature

### ❖ Nature AOP (Advance Online Publication)

- Released throughout weekdays
- 99 research articles
- Most found by DOI except for 6 by title in GS
- GS: 76% 4-5 days. Range from a few hours(4) to 8 days(6)
- Sci-Hub: 89% few hours. Range from a few hours(88) to 2 days(1)
- LibGen: 89% few hours. Range from a few hours(88) to 3 days(2)
- Google: 85% few hours. Range from a few hours(82) to 2 days(5)

# Currency of Information (Sep. 2016 to May 2017)

## ➤ Science

### ❖ Science New issues (research articles only)

- Released every Friday but in fact coming out Thursday night
- 68 articles only (many published as First Release)
- Most found by DOI except for 2 by title in GS
- GS: 87% 4-5 days. Range from a few hours(6) to 5 days(14)
- Sci-Hub: 78% few hours. Range from a few hours(53) to 24hrs(15)
- LibGen: 78% few hours. Range from a few hours(53) to 24hrs(15)
- Google: 85% few hours. Range from a few hours(57) to 24 hrs(10)



# Currency of Information (Sep. 2016 to May 2017)

## ➤ Science

### ❖ Science First Release

- Released throughout weekdays
- 124 research articles
- Most found by DOI except for 14 by title in GS
- GS: 84% 4-5 days. Range from a few hours(10) to 7 days(5)
- Sci-Hub: 60% 24 hours. Range from a few hours(49) to 24hrs(75)
- LibGen: 60% 24 hours. Range from a few hours(49) to 24hrs(75)
- Google: 73% few hours. Range from a few hours(87) to 24 hrs(33)



# WHAT IF?



WHY  
NOT?



**McGill**

Library  
Bibliothèque

Information ♦ Innovation ♦ Service